DOCUMENT RESUME

ED 261 097 TM 850 523

AUTHOR Abalos, Jose; And Others
TITLE Statistical Methods for Selecting Merit Schools.
PUB DATE Apr 85
NOTE 48p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Gains; Achievement Tests; Comparative Analysis; Cutting Scores; Elementary Education; Junior High Schools; *Merit Rating; *School Effectiveness; Socioeconomic Status; *Statistical Analysis; *Statistical Bias; Statistical Studies; Test Results
IDENTIFIERS Florida Merit School Program; *Palm Beach County Schools FL; Stanford Achievement Tests

ABSTRACT

This study investigated six statistical merit school selection methods using student scores on a nationally normed, standardized achievement test to identify merit schools. More specifically, its purpose was to select a method for the Palm Beach County School system which meets the Florida merit school program criterion of fairness in terms of socioeconomic status (SES). Stanford Achievement Test (SAT) 7 Reading Comprehension and Total Mathematics scores for grades 3, 4, and 5 were used in the elementary school analysis, and score for grades 7 and 8 in the middle/junior high school analysis. The methods of analysis investigated were (1) student-based regression; (2) school-based regression; (3) school-based multiple regression predicting mean gain with free lunch; (4) disaggregated percentile groups; (5) adjusting school mean scaled score gains by the differential performance of free and non-free lunch students; and (6) unadjusted mean scaled score gains. Four aggregation techniques combining indices across grades and subtests to yield a single school merit index (SMI) were presented and three methods of setting cutoff points were suggested. Results indicate that none of these merit school selection methods can be highly recommended. School-based regression analysis appeared the most acceptable for large school districts. Further research is recommended. (BS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED261097

Statistical Methods for Selecting Merit Schools

Jose Abalos, S. Jean Jolly, Robert Johnson
Palm Beach County School District

A paper presented at the Annual Meeting of the
American Educational Research Association
Chicago, Illinois: April 1985

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Jolly, S. Jean

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 850 523

INTRODUCTION

Recently there has been much interest in the use of test scores to select "effective" schools, primarily because of the development of a variety of district- and state-level school award programs. For the most part, awards are made to schools having high levels of student achievement or those showing improvement in achievement. The most commonly-used measure for classifying schools is an aggregate score on a nationally-normed, standardized achievement test (Kean, 1982). Methods for analyzing test scores are typically similar to those used in the effective schools research.

The simplest methods involve ranking schools by mean score or by proportion of students exceeding some standard. These absolute measures of instructional outcomes are, as a rule, inappropriate indices of school merit or effectiveness due to their bias in favor of schools populated by students with above-average socioeconomic status (SES) (Linn, 1983; Rowan, Bassert & Dwyer, 1983). The common practice of ranking schools within a state or a school district favors high SES schools, since the correlation between aggregated units of SES and academic achievement is .73 (White, 1983). Another common practice is to evaluate schools in terms of the proportion of students meeting or exceeding a standard related to national norms. For example, one of the criteria used in making school incentive awards in the Vanderbilt University (1984) school improvement project is a score at or above the national average. This method, similar to those used in the early effective schools research (Weber, 1971), is also highly biased in favor of high SES schools.

A second type of commonly used procedure is trend analysis, i.e., change in school performance across time. Gain scores produced from trend analysis are

generally inadequate measures of school merit or effectiveness (Frechtling, 1982; Linn, 1983; Rowan & Denk, 1983). Comparison of average scores at a particular grade level from one year to the next ignores possible changes in composition of a school's student body. Gain scores can be affected by changes in school demographics, making the method biased against schools growing in low SES students (Rowan & Denk, 1983). Moreover, schools do not necessarily evidence stable increases or decreases from one year to the next. For example, in a study of the scores of small, medium and large school districts on the Florida assessment test, Jolly (1982) found that small group averages were highly unstable from year to year.

Trend analysis can also be used with cohort groups when it is possible to match student scores from one year to the next. Although cohort analysis is an improvement over grade-level gain score analysis, the regression effect may bias results of cohort mean gain score analysis (Linn, 1983). An initially low-scoring cohort will tend to show positive gain, whereas an initially high-scoring group will tend to decline in achievement.

When computer facilities and technical expertise are available, the method most commonly used is regression analysis. Regression analysis of school mean achievement scores, controlling for student body SES factors, has become an increasingly popular method of producing meritorious or effectiveness indices (Purkey & Smith, 1982). Based on the regression equation, an expected score is calculated for each school and subtracted from the actual mean to yield a residual score, which is used as the index of effectiveness or merit. In studies using the student as the unit of analysis (Broward County School System, 1984; Frechtling, 1982; Kippel, 1981;

Marco, 1974; Soder, Sanders, Ramey & Matthews, 1981), the primary problem is the lack of control for group SES. White (1983) found that the correlation between academic achievement and SES was only .22 when the individual was used as the unit of analysis. For this reason the preferred strategy has been to employ regression analysis using school mean achievement scores, rather than individual student scores (Purkey & Smith, 1982). The school posttest mean is regressed on the school pretest mean, and a school residual score is calculated (Frechtling, 1982; Linn, 1983; Marco, 1974; Peng, Owings & Fetters, 1982; and Rowan & Denk, 1983). The major problem with this approach is the small number of schools typically available for developing the equation, which results in highly unstable coefficients. Whether the unit of analysis is the individual or the school, regression to the mean causes the residuals to be unstable from year to year, which, in turn, yields unreliable school rankings. Indeed, Rowan and Denk (1983) maintain that the residual approach cannot be improved sufficiently to justify its use.

In nearly all cases, indices of merit or effectiveness are based on average scores, whether the index be a mean gain or a mean residual. In their review and analysis of effective schools research, Purkey and Smith (1983) found that none of the studies examined achievement levels of different groups of students within the school. Yet use of the average ignores possible differences among subgroups of children (Cuban, 1983; Linn, 1983). Measures of effectiveness can vary widely for different types of students. Clauset and Gaynor (1982) argued that achievement data must be disaggregated to focus on the achievement patterns of different groups of students over time. In one such study, Marco (1974) selected reference points to represent low-, middle- and high-scoring students, and calculated three indices per school, using

the within-school regression model. The school effectiveness indices were the regression estimates of the mean posttest scores at the three reference points. He found that indices for initially low- and high-scoring students appeared to yield unique information. The correlation for these indices was .32, raising doubts about the utility of a single index as a school effectiveness measure. The Connecticut school improvement project (Pecheone & Shoemaker, 1984) uses an effectiveness index obtained by a disaggregation procedure. A frequency count is made of low- and middle-income students in each decile, and a t-test of proportions is used to assess growth changes in each of the ten equal units of the percentile scale for each school. Additionally, Linn (1983) suggests the use of different regression equations for high- and low-scoring students, and Clauset and Frederick (1984) suggest trend analysis using cohort gain of high- and low-SES students over a two- or three-year period.

Despite their obvious technical flaws, it appears that the demand for merit or effectiveness indices will continue to increase. For example, both California and Florida now have school awards programs in which test scores are a major factor. Additionally, numerous school districts have, or are in the process of developing, school awards programs. Although many of the methods tend to be highly correlated, they differ sufficiently enough from one another and in their relation to socioeconomic variables that they cannot be used interchangeably (Marco, 1974).

The purpose of this study was to investigate various merit school selection methods and related issues. More specifically, the impetus for the study was to identify a method which would satisfy the major Florida merit schools program criterion requiring that the method be fair and non-biased in terms of SES.

METHODOLOGY

Setting and Data Sources

The Palm Beach County school system has over 76,000 students housed in 100 schools, ranging in student composition from high minority/low income to low minority/high income. Student performance on the Stanford Achievement Test, Seventh Edition (SAT/7), varies from schools which score well above the national average to those scoring well below the norm. The county school district maintains a data file that includes information on each student such as race, sex, federal lunch eligibility, bilingual and gifted status, and retention from previous year. In addition, the SAT/7 scores for grades 2 - 9 are maintained in computerized data files.

In this study, 1983 and 1984 SAT/7 Reading Comprehension and Total Mathematics scores for grades 3, 4 and 5 were used in the elementary school analysis, and scores for grades 7 and 8 were used in the middle/junior high school analysis.

Procedures

This section of the paper delineates the six merit school selection methods under investigation in the study. In addition, this section contains descriptions of four techniques of aggregating merit measures across grades and subtests. Lastly, procedures for comparing selection methods are described.

Selection Methods

Past experience in the district with ranking schools using current year's test results or by using year-to-year grade level comparisons had shown these methods to be unacceptable for reasons well supported in the literature. Consequently, all methods used in the present investigation were based on cohort groups.

The first method investigated was student-based regression analysis (Method I). Students' current scaled scores were used as criterion, and prior achievement and SES/demographic variables were used as predictors. The SES/demographic variables used in this method included federal lunch eligibility, participation in bilingual and gifted programs, race, sex, age, and retention status. Residuals were calculated for each student and averaged across students within a school for a given subtest and grade level to generate a school mean residual for that grade and subtest. The school mean residuals were then aggregated across grades and subtests to yield a school merit index (SMI), and schools were ranked based on their SMIs.

The second method investigated was school-based regression analysis (Method II). The mean of the current year's scaled score for a given subtest and grade level was used as criterion, and prior year's mean achievement and percentage of students participating in the free federal lunch program were used as predictors. The number of predictors was limited to two due to the small number of schools in the study (60 elementary schools and 18 middle/junior high schools). School residuals

were computed by subtracting predicted from actual school mean achievement scores for every grade and subtest. These residuals were then aggregated across subtests and grade levels to yield an SMI for the school, and schools were ranked based on this SMI.

The third method investigated was a variation of the school-based regression analysis (Method III). The mean scaled score gain of students (current minus prior achievement) was used as criterion, and the percentage of students participating in the federal lunch program was used as predictor. The predicted mean gain score was subtracted from the school's actual mean gain score to generate a residual for a given subtest and grade level. The school residuals were then aggregated across subtests and grade levels to yield an SMI for the school, which was used as the basis for ranking the schools.

The fourth method investigated involved disaggregating the achievement data (Method IV). The total district population in each grade level was divided into ten percentile groups (i.e., 1st to 10th, 11th to 20th, etc.) using the students' prior year score on the Vocabulary subtest of the SAT/7. The Vocabulary subtest was used as the basis for grouping students due to the unavailability of a measure of school ability. The mean NCE gain in reading and mathematics of each percentile group within a school was then computed and compared to the mean gain of the district for the corresponding percentile group. The differences in gain (school minus district) for each group were then aggregated across grades and subtests to generate a composite index for each percentile group within the school. A final aggregation was made by

combining across the ten groups to yield an overall SMI which was used as the basis for ranking the schools.

The fifth method investigated involved adjusting school mean scaled score gains by the differential performance of free and non-free lunch students in the total county (Method V). In this method, the countywide difference in gain of students who participated in the free federal lunch program and those who did not was computed by subtracting their respective mean gain scores at each grade level and subtest. This differential index was used to adjust the mean gain achievement score of each school by multiplying the index by the percentage of students in the school who participated in the free federal lunch program and adding the product to the observed school gain. No adjustment was made at grade levels and subtests where the difference was found to be statistically insignificant. Adjusted mean gains were then aggregated across grades and subtests to yield an SMI which was used as the basis for ranking schools.

The last method investigated used the unadjusted mean scaled score gain of schools (Method VI). The school mean gain for a given grade level and subtest was computed by averaging the gain (current minus previous year's achievement) of each of the students in that particular grade level and subtest. These school mean gains were then aggregated across grades and subtests to yield an SMI which was used as the basis for ranking the schools.

Aggregation Techniques

Each of the six selection methods investigated generated an index for every grade level and subtest. Six indices (grades 3 to 5 for reading and math) were generated for elementary schools. Four indices (grades 7 and 8 for reading and mat.) were generated for middle/junior high schools. Four aggregation techniques combining indices across grades and subtests to yield a single SMI were investigated.

The first aggregation technique investigated was a significance test of the mean index of each school compared with the mean index of the district for a given grade and subtest. The mean index of the students in each school was compared with the mean index of the total district population to determine if the school measure was significantly higher, no different, or significantly lower than the performance of the total group of students. The t-test was used to compare these means at the .05 level of significance. A plus one (+1), zero (0), or minus one (-1) was assigned to each measure depending on whether the school mean was significantly higher, no different, or significantly lower than the district mean. These significant difference indicators were then averaged across grades and subtests to generate an SMI.

The second aggregation technique investigated was a variation of the significance test discussed above. Instead of assigning significant difference indicators (+1 ,0, and -1), however, the computed t-values were averaged across grades and subtests to yield a mean t-value. This mean t-value was used as the SMI, and schools were ranked on this basis.

The third aggregation technique investigated involved standardizing the measures and then averaging these standardized measures across grades and subtests to yield an SMI for each school.

The fourth aggregation technique, which was only used in conjunction with Method IV, involved averaging the NCE gain differences for each percentile group across grades and subtests, and finally across the ten percentile groups for each school to generate an SMI.

Comparison of Selection Methods

The resulting indices yielded by each selection method before aggregation were correlated with the SES/demographic variables across grades and subtests to evaluate the extent to which the methods eliminated bias. In addition, the indices generated after aggregation (SMIs) were correlated with federal lunch and race to determine whether the selection methods eliminated bias in the final aggregation. Lastly, the SMIs generated by the six selection methods were correlated with each other to determine the degree of relationship among the methods.

RESULTS

This section of the paper contains the results of applying the six selection methods to district data. Secondly, the results of aggregating the measures across grades and subtests using the four aggregation techniques are presented. Lastly, the results of correlating the indices generated by the selection methods with each other and with the SES/demographic variables before and after aggregation are reported.

Selection Methods

For Method I (student-based regression) the standardized regression weights across grades of the various predictor variables for reading and mathematics are shown in Table 1. Also in Table 1 are the number of students (N) for each grade level, the proportion of variance explained (R^2) by the combined predictor variables, and the standard deviation of the residuals (Residual SD) in scaled score points. It can be seen that previous achievement weighted more heavily (standardized regression weights ranging from .61 to .90) than any of the SES/demographic variables. Of the SES/demographic predictors, race, gifted and age weighted more than the rest. The resulting regression weights of the student regression analysis were very consistent across grade levels and subjects, with the only exception found in the slight variations of the weights of some of the SES/demographic variables.

The standardized regression weights for Method II (school-based regression predicting current achievement with previous achievement and federal lunch percentage) are shown in Table 2. The first two columns show the grade level and

Table 1

Results of Student-based Regression Analysis (Method I)
Across Grades and Subtests

READING COMPREHENSION

Grade	N	Previous Achiev.	Fed. Lunch	Race	Gifted	Age	Sex	Biling.	Retention Status	R^2	Residual SD
3	3280	.60		-.12	.21	-.14				.54	27.29
4	3290	.64		-.11	.10	-.08	.02		.05	.56	32.11
5	3700	.61	-.05	-.16	.04	-.12		.04		.60	30.93
7	4600	.66	-.06	-.06	.09	-.07		.04	.03	.60	25.78
8	4600	.73	-.05	-.08	.08	-.05	.02		.04	.69	26.69

TOTAL MATHEMATICS

Grade	N	Previous Achiev.	Fed. Lunch	Race	Gifted	Age	Sex	Biling.	Retention Status	R^2	Residual SD
3	3280	.64				-.13	-.14			.50	22.64
4	3290	.71	-.03	-.07	.06	-.10	.03	.04	.06	.63	23.77
5	3700	.75	-.04	-.07	.07	-.08		.04		.73	20.21
7	4600	.76	-.07	-.04	.13	-.04		.02		.76	19.06
8	4600	.87	-.04	-.04	.02	-.03			-.01	.83	17.55

Table 2

Results of School-based Regression Analysis
(Method II)
Across Grades and Subtests

READING COMPREHENSION

Grade	N	----Weights---- Previous Achiev.	Free Lunch	R^2	Residual SD
3	58	.92		.85	6.2
4	58	.71	-.22	.81	8.8
5	58	.67	-.32	.92	6.3
7	17	.80	-.19	.97	2.7
8	17	.57	-.42	.94	4.7

TOTAL MATHEMATICS

Grade	N	----Weights---- Previous Achiev.	Free Lunch	R^2	Residual SD
3	58	.41	-.53	.73	7.4
4	58	.63	-.29	.77	7.9
5	58	.63	-.34	.87	6.2
7	17	.82	-.17	.93	3.9
8	17	.97		.95	3.7

the number of schools at that grade. The next four columns show the standardized regression weights for previous achievement and free lunch, the variance explained (R^2) by the combined predictors, and the standard deviation of the residuals in scaled score points. It can be seen in Table 2 that weights for previous achievement varied from .97 to .41, and weights for free lunch varied from -.53 to zero. Note, however, the consistently high proportion of variance accounted for by the combination of previous achievement and federal lunch, ranging from .73 to .97.

The standardized regression weights for Method III (school-based multiple regression predicting mean gain with free lunch) are shown in Table 3. The first two columns show the grade and the number of schools at each grade level. The last four columns show the correlation of gain with free lunch on the school level, followed by the regression weight (if significant), the variance explained by the equation (R^2), and the standard deviation of the residual in scaled scores. Federal lunch did not weight significantly in some grade levels (e.g., grades 3 and 4) owing to the significantly low correlation between the criterion (gain) and the lone predictor (federal lunch). However, at some grades free lunch was a very strong predictor, weighing as high as -.54. For seventh grade Reading Comprehension, federal lunch showed a positive weight of .33.

For Method IV (disaggregated percentile groups), Table 4 shows the gains in normal-curve equivalent (NCE) units for each of the ten percentile groups across grades and subtests for the total district population. As shown in the table, district gains in the various percentile groups ranged from -3.9 to 9.2 NCE points. The gains

Table 3

Results of School Mean Gain Regression Analysis (Method III)
Across Grades and Subtests

READING COMPREHENSION

Grade	N	Correlation Gain/Fed. Lunch	Weight Fed. Lunch	R^2	Residual SD
3	58	.11			
4	58	-.21			
5	58	-.44	-.45	.21	6.5
7	17	.33	.33	.11	3.1
8	17	-.54	-.54	.30	5.0

TOTAL MATHEMATICS

Grade	N	Correlation Gain/Fed. Lunch	Weight Fed. Lunch	R^2	Residual SD
3	58	-.23			
4	58	-.19			
5	58	-.22			
7	17	-.34	-.34	.11	3.9
8	17	-.52	-.52	.27	3.6

Table 4

District Mean NCE Gains by Percentile Groupings (Method IV) Across Grades and Subtests

READING COMPREHENSION

Groups		1	2	3	4	5	6	7	8	9	10
Percentile Range		1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-99
Grade	3	4.4	2.7	2.8	-0.2	0.1	-0.7	-1.8	-1.7	-1.4	-2.8
	4	0.4	-0.4	-1.2	-2.1	-1.0	-1.5	-2.7	-0.2	1.0	-0.5
	5	-1.0	-0.5	1.1	2.4	4.2	4.8	3.0	4.7	4.2	1.0
	7	1.2	0.2	0.6	1.3	0.4	-0.5	-0.8	-0.5	-2.5	-2.4
	8	3.2	4.7	5.3	6.6	6.6	7.3	6.9	7.6	8.9	7.4

TOTAL MATHEMATICS

Groups		1	2	3	4	5	6	7	8	9	10
Percentile Range		1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-99
Grade	3	9.2	5.1	4.3	4.4	2.9	1.0	0.7	0.3	-2.4	-2.6
	4	0.5	0.8	-1.0	-0.1	0.6	-1.1	-1.0	-1.7	-1.5	-2.2
	5	3.8	1.2	0.8	0.6	0.4	1.0	0.6	1.5	2.1	1.8
	7	0.0	-1.9	-2.6	-2.2	-2.9	-3.6	-3.4	-3.9	-2.5	-1.5
	8	0.0	-0.3	0.4	1.1	1.5	0.4	1.4	1.2	-0.2	0.7

across the ten groups were highly variable, which formed the basis for the use of the disaggregation procedure.

The results of Method V (adjusting school mean gain by the differential performance of free lunch and non-free lunch students) are shown in Table 5. The number of students in each group together with their mean gains and the standard deviations of the gain in scaled scores are given in Table 5. The last column shows the mean difference between the gain scores of non-federal lunch and federal lunch students for the district. (A positive difference indicates that non-federal lunch students gained more than the federal lunch group.) These results showed much diversity in the mean gain differences of the two groups, with differences ranging from -3.4 to 7.7. It can be noted that in some grade levels federal lunch students gained more than non-federal lunch students.

Method VI (unadjusted mean gain) results across grades and subtests are given in Table 6. The last two columns show the mean gains in scaled scores and the standard deviations of these gains for the total population for each of the grade levels and subtests. The mean gains range from a low of 6.6 for seventh grade Reading Comprehension to a high of 41.5 for third grade Total Mathematics.

Aggregation Techniques

Table 7 shows the correlations among the three aggregation techniques across the five selection methods where they were used. The fourth aggregation technique was applied only to Method IV, thus the exclusion of this method from Table 7.

Table 5

Results of Adjusted Mean Gain (Method V)
Across Grades and Subtests

READING COMPREHENSION

Grade	Non-Federal Lunch N	Mean	SD	Federal Lunch N	Mean	SD	Mean Diff
3	2023	34.92	34.99	1219	36.29	34.44	-1.37
4	2069	14.89	35.84	1172	12.83	31.55	2.06*
5	2372	28.13	35.43	1305	23.63	32.01	4.50**
7	3198	5.56	30.67	1435	8.99	26.00	-3.43**
8	3381	26.06	28.01	1197	18.41	25.90	7.65**

TOTAL MATHEMATICS

Grade	Non-Federal Lunch N	Mean	SD	Federal Lunch N	Mean	SD	Mean Diff
3	1997	42.50	27.83	1200	39.89	28.07	2.61**
4	2064	21.73	25.51	1157	18.47	24.48	3.26**
5	2362	26.75	22.74	1289	25.88	20.52	0.87
7	3129	10.79	22.00	1398	7.85	19.61	2.94**
8	3306	22.52	18.37	1157	17.32	15.96	5.20**

* $p<.10$
** $p<.01$

Table 6

District Unadjusted Mean Gain (Method VI)
Across Grades and Subtests

READING COMPREHENSION

Grade	Mean Gain	SD
3	35.4	34.8
4	14.1	34.4
5	26.5	34.3
7	6.6	29.3
8	24.1	27.7

TOTAL MATHEMATICS

Grade	Mean Gain	SD
3	41.5	27.9
4	20.6	25.2
5	26.4	22.0
7	9.9	21.3
8	21.2	17.9

Table 7

Correlations Between Aggregation Techniques Across Selection Methods

ELEMENTARY SCHOOLS

Selection Method	I			II			III			V			VI	
Aggregation Technique	A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3
A1		.94	.91		.94	.91		.93	.89		.92	.88		.92	.85
A2			.97			.97			.97			.96			.89
A3															

MIDDLE/JUNIOR HIGH SCHOOLS

Selection Method	I			II			III			V			VI	
Aggregation Technique	A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3
A1		.95	.95		.93	.94		.94	.93		.92	.88		.95	.92
A2			.95			.96			.96			.96			.97
A3															

Overall, the aggregation techniques correlated highly with each other, with correlations ranging from .85 to .97 for elementary schools and from .88 to .97 for middle/junior high schools. Except for a few methods, the highest correlations were noted between aggregation techniques A2 (averaging significance values) and A3 (averaging standardized indices) and the lowest between A1 (significance difference indicators) and A3 (averaging standardized indices).

Correlations Among Selection Methods and with SES/Demographic Variables

Tables 8a and 8b show the correlations of the indices generated by each of the six selection methods with the SES/demographic variables before aggregation. Only correlation coefficients significant at the .05 probability level were noted.

It can be seen that Method VI (unadjusted mean gain) and Method I (student-based regression) had the strongest correlations with the SES/demographic variables, both in numbers and in magnitude. The correlations of Method VI with free federal lunch ranged from -.55 to .34 across the various grade levels. Note also the significant correlations of this method with race, with coefficients ranging from -.49 to .51. Method I showed significant correlations with free federal lunch and race in grades 3, 5 and 7, with correlations registering as high as -.45 with federal lunch and -.41 with race.

Method III (school gain regression), Method IV (PR Grouping) and Method V (adjusted mean gain) registered fewer significant correlations than Method I (student-based regression) and Method VI (unadjusted mean gain). Method V registered

Table 8a

Correlations of Pre-Aggregation Indices with SES/Demographic Variables Across Selection Methods

READING COMPREHENSION

METHOD

	I STUDENT REGRESSION					II SCHOOL REGRESSION					III SCHOOL GAIN REGRESSION				
Grade	3	4	5	7	8	3	4	5	7	8	3	4	5	7	8
SES/Demographic Variables															
Federal Lunch	.29	-.21	-.38	-.45	.	.	.	.	.	.	.	-.23	.	.	.
Race	.32	.	-.41	-.29	.	.	.	.	.	.	.	.	.	.	.
Gifted	-.70	.	.29	.44	.	.	.	.	.39	.	.	.23	.	.	.
Sex	.	.	.29	.	.	.	.	.	.25	.	.	.	.	.43	.
Age	.31	.	.	.	.	.	.	.	.	.	.	.	.23	.	.
Bilingual	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Retention Status	.25	.	-.41	-.58	.	.	.	-.25	-.48	.	.	.	.	-.34	.

	IV PR GROUPING					V MEAN GAIN ADJUSTED					VI MEAN GAIN UNADJUSTED				
Grade	3	4	5	7	8	3	4	5	7	8	3	4	5	7	8
SES/Demographic Variables															
Federal Lunch	.	.	-.28	.	.	.	-.23	-.30	.	.	.	-.23	-.44	.34	-.54
Race	.	.	-.30	.	.	.	.	-.37	.30	.	.	.	-.49	.51	.
Gifted	.24	.	.	.46	.42	.	.23	.	.	.	.	.23	.25	.	.
Sex	.	.	.	.	.	.	.	.26	.44	.	.	.	.27	.43	.
Age	.	.	.	.	.	.	.	.	.27	.	.	.	.	.46	-.54
Bilingual	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Retention Status	.	.	.	.	.	.	.	-.25	.30	.	.	.	.29	.	.

Table 8b

Correlations of Indices with SES/Demographic Variables Across Selection Methods

TOTAL MATHEMATICS

	METHOD														
	I STUDENT REGRESSION					II SCHOOL REGRESSION					III SCHOOL GAIN REGRESSION				
Grade	3	4	5	7	8	3	4	5	7	8	3	4	5	7	8
SES/Demographic Variables															
Federal Lunch	-.39	.	-.25	-.33	.	.	.	.	.	.	-.22	.	-.23	.	.
Race	-.28	.	.	-.24	.	.	.	.	.	.	.	.	.	.	.
Gifted	.35	.	.	.48	.	.33	.	.	.37	.	.23	.	.	.27	.
Sex	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Age	.	-.27	-.24	.	.	.	.	.	.	.	.	-.30	-.24	.	.
Bilingual	.	.	.	-.46	.	.	.	.	-.33	.	.	.	.	-.37	.
Retention Status	.	.	-.24	-.34	.	.	.	-.22	-.26	.	.	.	.	-.24	.

	IV PR GROUPING					V MEAN GAIN ADJUSTED					VI MEAN GAIN UNADJUSTED				
Grade	3	4	5	7	8	3	4	5	7	8	3	4	5	7	8
SES/Demographic Variables															
Federal Lunch	-.27	.	-.22	.	.	.	.	-.23	.	.	.22	.	-.23	-.33	-.55
Race	.	.	.	.	.	.	.	.	.	.	.	.	.	.24	-.46
Gifted	.25	.	.	.	.	.	.	.	.35	.	.23	.	.	.41	.
Sex	.22	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Age	.	-.30	.	.	.	.	-.23	-.24	.	.	.	-.30	-.24	.	-.50
Bilingual	.	.	.	.	.	.	.	.	-.45	.	.	.	.	-.51	.
Retention Status	.	.	.	.	.	.	.	.	-.30	.	.	.	.	-.34	.

some significant correlations with free federal lunch and race, especially in Reading Comprehension. Method IV registered some significant correlations with both free federal lunch and race for grade 5 in Reading Comprehension and with federal lunch for grades 3 and 5 in Total Mathematics. Method III did not register any significant correlation with race and only slight correlations with free federal lunch in grade 4 in Reading Comprehension (-.23) and in grades 3 and 5 in Total Mathematics (-.22 and -.23, respectively).

Method II (school-based regression) showed the least number of significant correlations with the SES/demographic variables. More importantly, it showed no significant correlations with either free federal lunch or race in any of the grade levels.

For the indices after aggregation (SMIs), the first part of Table 9 shows the correlations for elementary schools of these indices with free federal lunch and racial composition. Method VI (unadjusted mean gain) showed the highest post-aggregation correlation with federal lunch, followed by Methods IV, I, V, and III. The correlations of the indices with free federal lunch for Method VI (unadjusted mean gain) ranged from -.36 to -.43. Method IV (PR Grouping) registered a -.41 correlation; while the correlations for Method I (student-based regression) ranged from -.28 to -.37. Method V (mean gain adjusted) showed correlations of -.27 to -.34, and Method III (school gain regression) evidenced correlations of -.22 to -.31 Lastly, as in Table 8, Method II (school-based regression) registered the least correlation with federal lunch regardless of the aggregation technique used. Note that the correlations of the SMIs with free federal lunch for this method across the three aggregation approaches were

Table 9
Correlations Between Post-Aggregation Indices (SMIs) and Major SES/Demographic Variables

ELEMENTARY SCHOOLS

Selection Method	I			II			III			V			VI			IV
Aggregation Technique	A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3	A4
Federal Lunch	-.31	-.37	-.28	-.06	-.07	-.04	-.24	-.31	-.22	-.27	-.34	-.29	-.36	-.43	-.37	-.41
Race	-.17	-.26	-.19	-.02	-.05	-.05	-.18	-.25	-.18	-.20	-.27	-.23	-.29	-.35	-.31	-.36

MIDDLE/JUNIOR HIGH SCHOOLS

Selection Method	I			II			III			V			VI			IV
Aggregation Technique	A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3	A1	A2	A3	A4
Federal Lunch	-.33	-.30	-.22	-.12	-.06	-.02	-.11	-.03	-.05	-.29	-.19	-.07	-.46	-.40	-.29	-.40
Race	-.26	-.21	-.15	-.03	.00	-.06	-.06	.02	-.08	-.25	-.14	-.03	-.41	-.36	-.25	-.29

Note: Method I - Student-based Regression
II - School-based Regression
III - School Gain Regression
IV - Percentile Grouping
V - Adjusted Mean Gain
VI - Unadjusted Mean Gain

Aggregation Technique A1 - Significance Indicators
A2 - Mean Significance Values
A3 - Mean Standardized Indices
A4 - Mean NCE Gain Difference

very low, ranging from -.04 to -.07. Overall, a similar pattern was evident for the correlations of SMIs with race except that Method IV, instead of Method VI, registered the highest correlation.

For middle/junior high schools, the correlations of the SMIs with federal lunch and racial composition are shown in the bottom of Table 9. It can be seen that Methods VI and IV evidenced the highest correlations, followed by Methods I and V. Methods III and II showed the least correlation with federal lunch and racial composition.

The correlations of the SMIs of the six selection methods for the elementary and middle/junior high schools are shown in Table 10. It will be noted that the correlations among the six selection methods were highly positive with correlation coefficients ranging from .75 to .99 for the elementary schools, and .64 to .99 for the middle/junior high schools.

Table 10

Correlations Among SMIs of Selection Methods

ELEMENTARY SCHOOLS

SELECTION METHODS	I	II	III	IV	V	VI
I. Student-based regression		.80	.87	.84	.88	.86
II. School-based regression			.87	.75	.85	.82
III. School Gain regression				.91	.99	.98
IV. PR Grouping					.92	.93
V. Mean Gain Adjusted						.99
VI. Mean Gain Unadjusted						

MIDDLE/JUNIOR HIGH SCHOOLS

SELECTION METHODS	I	II	III	IV	V	VI
I. Student-based regression		.90	.88	.80	.90	.89
II. School-based regression			.99	.70	.99	.94
III. School Gain regression				.64	.99	.93
IV. PR Grouping					.68	.72
V. Mean Gain Adjusted						.97
VI. Mean Gain Unadjusted						

Note: Aggregated NCEs were used for Method IV. The rest of the selection methods were based on standardized residuals.

DISCUSSION

This section of the paper is divided into two major subsections: (1) issues relating to the selection methods and (2) issues relating to aggregation and setting cutoff points for selection.

Issues Relating to Selection Methods

Examination of the results of the different selection methods raised a number of issues and concerns which are discussed in this subsection of the paper. These issues include the appropriate unit of analysis, the effect of the number of observations, consistency of selection models across grades and subtests, and bias. Finally, the strengths and weaknesses of the various methods are discussed.

Unit of Analysis

The question of the appropriate unit of analysis in research studies has been raised and discussed repeatedly in the literature (Blair & Higgins, 1984; Fletcher, 1968; Hopkins, 1982; Raths, 1967). In using student-based regression analysis for evaluating schools, there are some concerns expressed about the violation of the assumption of independence of observations. Some authors note that if the final use of a model is to compare school performance, creating the model on another basis (such as students) is not appropriate. On the other hand, Hopkins (1982) contends that the "common recommendation to use group means where there may be nonindependence among observational units is unnecessary, unduly restrictive,

impoverishes the analysis, limits the questions that can be addressed in a study, and does not insure that the relevant independence assumption has been met."

Number of Observations

The issue of number of observations becomes critical in merit/effective school programs when applied on the district level because of the diversity of district size. The majority of school districts have fewer than a hundred schools for any given grade grouping configuration. If multiple regression analysis is to be utilized in the selection process, the number of observations available makes the question of whether to use student- or school-based regression crucial. School-based regressions, because of problems associated with limited N's, may yield unstable coefficients and be highly variable across grade levels and subtests, as was found in this study. Additionally, the number of predictors will be seriously limited, and resulting regression formulas may be highly influenced by outliers which may bias the outcome towards certain schools. In most cases, districts would have too few schools to warrant the use of school-based regression.

Student-based regression provides the luxury of a larger number of observations, which in turn results in more consistent regression coefficients as was shown in our results. It is quite evident that consistency of regression weights across grades and subtests is more pronounced in regression equations that are based on larger Ns. Additionally, because of the larger N, student-based regression allows the inclusion of more predictors. Moreover, the influence of outliers is not an issue as when the equation is generated by a small N.

Other methods, such as Method IV (percentile groupings) and Method V (adjusted mean gain), are relatively free from the problems associated with the number of observations since these methods are based on differential performance of various ability levels (Method IV), or of students of dichotomous SES groups (Method V).

Consistency of Selection Models Across Grades and Subtests

In the area of consistency of models, the regression coefficients generated for Method I (student-based regression) were found to be very uniform across grades and subtests. On the other hand, Method II (school-based regression) showed large differences in regression weights across grades and subtests, especially for federal lunch. Method III (school gain regression) showed the same variability, with federal lunch weighting from -.54 to .33. Neither of these school regressions exhibited any subetantial consistency across grades and subtests.

The question of consistency was not appropriate for Method IV (percentile groups), in that the method supposes different mean gain for each ability grouping. In similar manner, Method V (adjusted mean gain) and Method VI (unadjusted mean gain) showed considerable variation across grades, although this might be the result of test characteristics and student performance and not a question relative to the model.

Additionally, inconsistencies were noted across methods. For example, Method II (school-based regression) weights federal lunch in grade 7 Reading Comprehension at -.19, whereas the same factor was weighted at .33 in Method III (school gain

regression). For Total Mathematics at grade 8, federal lunch weighted at -.52 in Method III but not at all in Method II.

Bias

A major test of any method is to determine if it is insensitive to the SES variables which are considered important. The findings of this study indicated that using Method VI (unadjusted mean gain) for selecting merit schools would result in high SES/low minority schools having a greater probability of being selected than low SES/high minority schools. In short, using unadjusted mean gains in identifying meritorious schools results in a highly biased selection procedure.

Although Method I (student-based regression) yields very consistent regression coefficients across grades, it fails to eliminate the bias against low SES/high minority schools. Correlations in the range of .2 to .4 are common before aggregation, with similiar values observed after aggregation. This result might be explained by the fact that the relationship between academic performance and socioeconomic status is much lower on the individual level than on the aggregated level (see White, 1983).

For Method IV (percentile groupings), it was our expectation that by adjusting for the varying achievement of different student ability groups, most of the SES bias would be removed. This method exhibited some bias at the .2 to .3 level before aggregation, but surprisingly correlated very highly with federal lunch on the final aggregation.

Method V (adjusted mean gain) registered some significant correlations with free federal lunch and race, especially in Reading Comprehension. Apparently the adjustment of school gains by the differential performance of federal and non-federal lunch students in the total county was insufficient to remove all bias in terms of SES.

In comparison with the other selection methods, Method II (school-based regression) resulted in the least biased selection of schools, as evidenced by the non-significant correlations with the SES variables before and after aggregation. Method III (school gain regression) showed slightly more bias than Method II, although the inclusion of regression coefficients that were deemed non-significant on the school level would probably have reduced the correlation to the level of Method II.

The reason for the remaining bias of Methods IV (percentile groupings) and Method V (adjusted mean gain) was also considered. Whereas it is understandable that Method I (student-based regression) remained somewhat biased since the relationship of SES to achievement is weak on the student level, the bias remaining in Method IV and Method V was somewhat puzzling. Method V, adjusting mean gain of schools by the differential performance of free and non-free lunch students, should have removed virtually all bias. In the same vein, it was also expected that Method IV, by comparing the gains of low-scoring students with those of similiar ability, would remove most bias.

For Method V, a possible explanation for the bias left after the adjustment of mean gain by the difference in the performance of free/non-free lunch students could

be that the percentage of students participating in the federal free lunch program does not adequately measure the SES of the school. For example, a school with 90 percent free lunch students may really be lower in SES than of .90 times the mean of all the free lunch students in the district. In other words, many of the students in the total group of free lunch students may have higher income levels than those at that particular school.

The bias left in Method IV (percentile grouping) indicates that low-ability students in a low SES school tend to gain less than students of equally low ability in the county as a whole. It could be that a high percentage of low SES students in a school creates an additional factor which affects achievement.

In summary, therefore, it appears that Method II (school-based regression) yields the least biased results. The next preferable method would be Method III (school gain regression). Method I (student-based regression), Method V (adjusted mean gain), and Method IV (percentile grouping) all show biased results in terms of SES. Ranking schools based on Method VI (unadjusted mean gain) results in the most biased selection of schools.

Strengths and Weaknesses of Selection Methods

The apparent strengths and weaknesses of the various merit school selection methods are as follows:

Method I (student-based regression) has the advantage of the large N of students, tending to yield regression coefficients that are consistent across grade levels and subtests. It also allows for the use of more predictor variables and is not greatly influenced by outliers. However, this method has the distinct disadvantage of the weak relationship of SES variables to academic performance at the student level, resulting in bias in selection due to inadequate adjustment for SES.

Method II (school-based regression) has the advantage of most closely fitting the school data, removing practically all apparent bias on the variable being used for the adjustment. However, the small N of schools may yield unstable and highly varying regression equations, being overly sensitive to the specific values of certain schools, and subject to the effects of outliers. The relative weights of prior achievement and federal lunch were also highly variable. Finally, in many cases insufficient number of schools may exist in a district to enable the use of this method.

Method III (school gain regression) shares the same basic advantages and disadvantages as Method II (school-based regression). This method has the additional advantage of eliminating the differential weighting of previous achievement. However,

some bias still remained due to the non-significant correlations between gain and federal lunch which produced no adjustment when generating the regression equation.

Method IV (percentile groupings) has the major advantage of not utilizing regression analysis thereby eliminating most of the problems associated with regression, e.g., small Ns, unit of analysis, inconsistency of regression coefficients across grades and subtests. Although this method eliminated most of the bias prior to aggregation, it failed to do so in the final aggregation, and consequently remained biased.

Method V (adjusted mean gain) also has the advantage of adjusting for SES bias without the use of regression and its subsequent problems. However, in the final analysis it still appears to show moderate SES bias.

Method VI (unadjusted mean gain) has the advantage of being the simplest model. Having no adjustment for SES, however, this method yields the most biased results.

Issues Relating to Aggregation and Setting Cutoff Scores

This section discusses the problems related to aggregating results of selection methods across grades and subjects. A brief discussion of approaches for setting a cutoff for merit selection is then presented.

Aggregation Techniques

Combining indices across grade levels and subjects for the various selection methods was also an important consideration in this study. First, in combining indices the problem of adjusting for the different variability at different grade levels exists. The second problem is determining what amount of difference in the final index represents a real difference in the original observations. To address these problems, four techniques were used: a significance test (+1, 0, -1) of the school index (A1); averaging the significance tests (A2); standardizing the index and averaging (A3); and generating an index based on NCE's and averaging (A4).

The significance testing of the school index compared to that of the total county addresses the question of significant difference in the original observations. In the first variation of this method (A1), the "significant difference" indicators (+1, 0, -1) at least provide some indication of real difference. However, collapsing the t-test itself into three indicators results in the loss of information, and may create a problem in the case of school observations very close to the significance cut-off.

The second technique, averaging the results of the significance tests themselves (A2), solves the problems of loss of information and the separation of observations near the significance cutoff. However, this technique may raise some questions of statistical propriety. Additionally, the "average" significance test presents problems of interpretation.

The third technique, averaging of standardized residuals (A3), is rather straightforward and solves the problems of differing variance at different grades. However, the results of averaging, being means of standardized measures, are very difficult to interpret since the original metric has been lost. The interpretation of any particular difference, for example, is unclear in that the size of the difference relative to the proportion of error in the measurement is unknown.

The last aggregation technique, averaging of NCE indices (A4), was only attempted for Method IV (percentile groupings). This technique has the advantage of preserving the original metric, yielding a final index that is interpretable. If the standard deviations of the original NCE indexes is relatively constant from grade to grade, the final index should be free from problems of unequal weighting of grade levels.

Evaluation of Cutoff Methods

An important aspect of merit/effective school programs that is often taken for granted is the setting of a cutoff point to decide which schools deserve to be labeled "meritorious" or "effective". The following ways of setting cutoff points are suggested: (1) selecting a given percentage from the top of the distribution, (2) selecting a cutoff above one standard deviation of the distribution of the final indices, and (3) attempting to set an interpretable cutoff based on the properties of the index used. The problems involved in setting a cutoff are determining that the selection point involves a real difference from the mean of the schools, and avoiding the difficulty of separating schools at the cutoff which possess no valid difference.

The first approach (selecting a given percentage of schools) is certainly defensible and may easily be used. However, the question of significance of difference is ignored, as well as the question of separating indistinguishable schools at the cutoff point. The second approach (above one standard deviation of the indices) perhaps addresses the concern of significant difference, although only the difference among the schools (not absolute real difference) is evaluated. The concern of separating indistinguishable schools at the cutoff is still present. Finally, the third approach (attempting to set a "meaningful cutoff" based on the properties of the index being used) is possible in the cases of the techniques involving average significance tests or using average NCE gain. In the case of the significance indicators (+1, 0, -1), the final index represents the frequency of significantly differing performance for that school. Interpretation is rather easy, and since the final index will tend to be grouped (into a finite number of possible values) the problem of separating indistinguishable schools at the cutoff is minimized. For the NCE index, setting a meaningful cutoff may still be a problem, yet the preservation of the metric may permit specifying a minimum difference beyond which no distinction would be made.

RECOMMENDATIONS

The results of this study suggest that none of the preceding merit school selection methods can be highly recommended for adoption by school districts considering implemention of a merit school program. Of the six methods evaluated, school-based regression analysis appears to be the most acceptable, provided a large number of schools is involved in the program; otherwise, the stability of the regression equations becomes questionable. Unfortunately, it is unlikely that most districts would contain sufficient schools to warrant use of this method.

Clearly, a simple ranking of schools on their unadjusted mean achievement gain is not advisable. It results in a highly biased, and therefore unfair, school selection procedure. If, however, a district is forced to select a method, it might consider the adjusted mean gain (Method V). This method is not as highly biased as the unadjusted mean gain (Method VI), and it is relatively simple to apply and to explain.

The discrepancy noted in the pre-aggregation and post-aggregation results of Method IV (PR grouping) is difficult to explain. It was expected that this method would be the best technique for eliminating most of the problems associated with regression analysis, and we still maintain that disaggregated group procedures hold promise. It is, therefore, recommended that further studies of disaggregation methods be undertaken.

It is recommended that longitudinal studies of merit/effective school selection methods be conducted to test model stability across several years. Lastly, the important issue of establishing significant or meaningful cutoff points to distinguish between meritorious/effective schools and non-meritorious/non-effective schools warrants further empirical investigation.

BIBLIOGRAPHY

Blair, R. C., & Higgins, J. J. (1984). The unit of analysis: Group means versus individual observations. Florida Journal of Educational Research, 26(1), 5-20.

Broward County School System. (1984). Regression analyses of test scores and merit pay: Some statistical byproducts. Ft. Lauderdale, FL.

Clauset, K. H., Jr., & Gaynor, A. K. (1982). A systems perspective on effective schools. Educational Leadership, 40(3), 54-59.

Clauset, K., & Frederick, J. (1984). How to measure your school's effectiveness: A citizen's notebook on school effectiveness. Massachusetts: Institute for Responsive Education.

Cuban, L. (1983). Effective schools: A friendly but cautionary note. Phi Delta Kappan, 64(10), 695-696.

Fletcher, H. J. (1968). Possible interpretive problems in analyses using group means as the experimental unit. Psychological Bulletin, 69(3), 157-160.

Frechtling, J. A. (1982). The measurement of effectiveness. Paper presented at ERS/EN Annual Meeting, Baltimore, MD.

Hopkins, K. (1982). The unit of analysis: Group means versus individual observations. American Educational Research Journal, 19(1), 5-18.

Jolly, J. (1982, November). Relationship of district size and state assessment test scores. Paper presented at the meeting of the Florida Educational Research Association, Orlando, FL.

Kean, M. H. (1982). Issues in identifying effective schools. Paper presented at the 12th annual conference on Large Scale Assessment, Boulder, CO.

Kippel, G. (1981). Identifying exceptional schools. New Directions for Program Evaluation, 11, 83-100.

Linn, R. L. (1983). Measuring school effectiveness: How achievement data can and cannot be used. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Marco, G. L. (1974). A comparison of selected school effectiveness measures based on longitudinal data. Journal of Educational Measurement, 11(4), 225-234.

Pecheone, R. L., & Shoemaker, J. (1984). Tracking the footprints of school improvement: A study of Connecticut's school effectiveness project. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Peng, S. S., Owings, J. A., & Fetters, W. B. (1982). Effective high schools: What are their attributes? Paper presented at the annual meeting of the American Educational Research Association, Washington, D. C.

Purkey, S. C., & Smith, M. S. (1982). Too soon to cheer? Synthesis of research on effective schools. Educational Leadership, 40(3), 64-69.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. The Elementary School Journal, 83(4),

Rath, J. (1967). The appropriate experimental unit. Educational Leadership, 25(3), 263-266.

Rowan, B., Bassert, S.T., & Dwyer, D. C. (1983). Research on effective schools: A cautionary note. Educational Researcher, 12(4), 24-31.

Rowan, B., & Denk, C. E. (1983). Modelling the academic performance of schools using longitudinal data: An analysis of school effectiveness measures and school and principal effects on school-level achievement. San Francisco: Far West Laboratory.

Soder, J., Sanders, G., Ramey, M., & Matthews, T. (1981). Educational effectiveness and student academic gains. (ERIC Document Reproduction Service No. ED 209 366).

Vanderbilt Institute for Public Policy Studies (1984). Questions and answers about the school incentives project. Nashville, TN.

Weber, G. (1971). Inner city children can be taught to read: Four successful schools (Occasional Papers No. 18). Washington, D. C.: Council for Public Education.

White, K. R. (1983). The relation between socioeconomic status and academic achievement. In R. J. Light (Ed.), Evaluation Studies Review Annual, 8, 602-622.